DOCUMENT RESUME

ED 072 478 CS 500 125

AUTHOR Bochner, Arthur P.
TITLE Task and Instrumentation Variables as Factors Jeopardizing the Validity of Published Group Communication Research, 1970-1971.
PUB DATE Dec 72
NOTE 20p.; Paper presented at the Annual Meeting of the Speech Communication Assn. (58th, Chicago, December 27-30, 1972)

EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS *Communication (Thought Transfer); Evaluation Criteria; *Group Dynamics; Reliability; Researchers; *Research Methodology; *Research Problems; *Research Reviews (Publications); Standards; Validity

ABSTRACT

The validity and reliability of small group research published between 1970 and 1971 is examined in this paper. In response to the small group research position which gives precedence to theory over method, the author counters that placing measurement in a secondary position increases the danger of accepting claims of experiments which contain excessive error. A representative sampling of eight studies published in "Speech Monographs" between January 1970 and December 1971 were critically examined. The author found (1) that there is considerable agreement about what should be measured, but considerable disagreement about how to measure; (2) that neglect of task stimuli seriously comprised the internal validity of a number of studies; and (3) that accuracy and completeness of reliability was not a crucial factor in manuscript publication. The author argues for more stringent methodological standards in the publication of research and discusses suggestions for appropriate utilizations of tasks and complete and accurate assessment of reliability. (Author/LG)

U S DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY

ED 072478

Task and Instrumentation Variables as Factors Jeopardizing the Validity of Published Group Communication Research, 1970-1971

Arthur P. Bochner

Cleveland State University

PERMISSION TO REPRODUCE THIS COPYRIGHTED MATERIAL HAS BEEN GRANTED BY
Arthur P. Bochner

TO ERIC AND ORGANIZATIONS OPERATING UNDER AGREEMENTS WITH THE U.S. OFFICE OF EDUCATION. FURTHER REPRODUCTION OUTSIDE THE ERIC SYSTEM REQUIRES PERMISSION OF THE COPYRIGHT OWNER

Presented at the Speech Communication Association
National Convention, Chicago, December 1972

CS 500 125

Task and Instrumentation Variables as Factors Jeopardizing the Validity of Published Group Communication Research, 1970-1971

The basic objective of empirical research in speech communication is to acquire knowledge about the communication process. Researchers who are actively engaged in the pursuit of this task are often divided on the question of how best to achieve it. Is it better to concentrate research effort on generating theory and hypotheses or should problems of method and measurement precede experimentation? Recent critics of the small group field have advised researchers to give precedence to theory over method. This paper is written as a response to that position and is based on the following thesis: Placing measurement in a secondary role increases the danger of accepting the truth claims of experiments which contain excessive error. Since the internal validity of an experiment demands an unbiased estimate of experimental effects, excessive error diminishes the confidence that may be placed in experimental results.

McGrath and Altman's 1966 treatise[1] on the sorry state of small group research initiated an era of introspection and criticism for small group researchers. In speech communication, evaluations of small group studies have been reported by Bormann,[2] Gouran,[3] Mortenson,[4] Larson,[5] and Fisher.[6] Each of these critics has found that small group research suffers from a lack of accumulated knowledge, is characterized by conceptual ambiguity, is marred by poorly conceived research, and is devoid of theory. While they unanimously condemn what is studied, these writers find little wrong with the methods available to study it. A few of the critics find "social psychology methodology" inappropriate to a "communication" emphasis,[7] but for the most part they agree that the methods used are rigorous. Bormann refers to this as "the paradox of small group research": sophisticated

methodology coupled with barren results.[8] Mortenson believes that the lack of a communication emphasis in small group research can not be attributed to unsuitable methodology.[9] Gouran suggests that statistical procedures often interfere with the determination of significant research questions.[10] And McGrath and Altman contend that there has been an "overemphasis on procedures and an underemphasis on theory."[11]

These reports have served to place measurement and methodology in a secondary position. The small group researcher has been advised to "allow back into the formative stages of research the forces of intuition, allegiance to problems, and projection of the researchers subjective insights. Then, let methodological rigor enter..."[12]

These collective insights and recommendations give the impression that methodology is a stumbling block rather than an aid to research. Byrne has alluded to the risk involved in accepting such advice by reversing the viewpoint: "The primary advantage of giving precedence to measurement is that one's experimental results are not based on the shifting sands of error variance."[13]

This paper examines published small group research in speech communication during 1970 and 1971 for the purpose of determining how much confidence can be placed in the reported results. It is especially concerned with locating uncontrolled or unmeasured sources of error variance.

Tasks as a Source of Variance

The decision was made to limit the search for uncontrolled sources of variance to one class of variables. The critiques by Bormann, Gouran, and Mortenson emphasized the need for a communication orientation in small group research. Therefore, it was assumed that a majority of the published research would either measure or manipulate communication behavior. With this in

mind, the following question was asked: Is there a class of variables which contributes significantly to variation in group communication behavior and is often left uncontrolled and unmeasured?

The interactionist position in social psychology states that two classes of variables affect behavior: the individual characteristics of the subject and the environmental (physical and social) characteristics of the situation.[14] Put another way, who a person is and what he is asked to do (or where he is) interact to account for most of the variation in his behavior. Since vital individual traits of group participants are usually either controlled (e.g. all males, all females, all college sophomores, etc.) or manipulated (e.g. group composition studies), it seemed more desirable to investigate the situational variables as a possible source of error variance. For the purpose of this paper, what a group was asked to do was defined as the task it was assigned.

A review of the small group literature provided ample justification for this decision. A number of small group researchers have censured the common practice of disregarding systematic task analysis.[15] Typical of the criticism is Shaw's allegation that the task variable is among the most neglected in social science research[16] and Weik's warning that tasks probably contain more uncontrolled sources of variance than any other components of experiments.[17]

The Sample

Eight studies[20] appearing in Speech Monographs between January, 1970 and December, 1971, were chosen as a representative sample of small group communication research. Each of these studies collected data for the express purpose of advancing knowledge about small group communication. In each study, the data was collected under laboratory conditions, although

only three reports actually manipulated independent variables.

The evaluation of the published research was guided by three questions:

(1) Does the published research indicate any areas of general agreement about what should be measured?

(2) Does the published research clearly report an unbiased estimate of experimental effects?

(3) Are the dependent variables reliably measured?

Investigation of the first question led to the comparison of the stated objectives or purposes of the study. The second question revolved directly around how tasks were utilized in the sampled studies. This included how the task was defined, whether it was controlled or manipulated, and its potential for contributing to error variance. The third question necessitated the search for reliability estimates of the dependent variables.

In order to facilitate a comparison of the eight articles a flow sheet was prepared. The flow sheet included the name of the author, the issue in which the article appeared, the stated purpose of the study, a description of the task assigned to the participating groups, and the author's stated conclusions. This flow sheet is presented in Table I.

- -

Insert Table I about here

- -

Convergence of Purpose and Divergence of Method

Seven of the eight studies measured either part or all of the verbal communication among group participants. Only the study by Burgoon was limited to the exclusive use of a self-report measure of participant reaction.

Three of the studies were designed to find the patterns (Bostrom and Fisher) or structure (Stech) of verbal acts in group discussion. The three

differed drastically in approach and completely in measurement of the verbal acts. Three other studies (two by Leathers and one by Larson) were similar; each was designed to compare the verbal behavior under one set of conditions with the verbal behavior under another set of conditions. Once again, both investigators used different dependent variable measures. In fact, there were seven different measures of verbal behavior used in these seven studies. Five of the eight studies employed instruments developed by the authors (Fisher, Stech, two by Leathers, and McCroskey, et.al.). None of the instruments had previously appeared in published research by other investigators. Since only the McCroskey and Wright instrument was factor analytically derived, there appears to be little chance that the others will receive any general use.

The question asked was, does the published research indicate any areas of general agreement about what should be measured? These eight studies suggest that there is considerable agreement about what should be measured, verbal communication behavior, but considerable disagreement about how to measure it. For this reason, there appears to be little comparability between studies.

Utilization of Task Variables

What is the liklihood that the tasks presented to the groups in these studies produced biased estimates of experimental effects? To answer this question, the task information supplied in each study was applied to three related questions: (1) How similar were the task assignments across the eight studies? (2) What types of tasks were used? (3) How were these tasks used?

Among the eight studies there was little similarity in what the groups were asked to do. Only Stech used a task which had previously appeared in

the literature, the Goldberg Music Ranking Task. Leathers' two studies and the one by McCroskey and Wright employed "campus-oriented" problems. Burgoon required his groups to reach consensus on a national proposition. Bostrom assigned a different topic to each of thirty groups, but the topics were not provided in the article. Larson gave "similar task assignments" to his groups, but the only information presented was that following discussion, each group had to "present a written and oral report focusing on a problem which concerned them". Fisher reported that the ten groups in his study discussed "decision-making tasks". In the Fisher study, two criteria for selection of the groups were given, they had to reach consensus, and they did, but no information about the specific dimensions of the tasks was provided. Fisher tells us that "the time required to solve the decision-making tasks varied from 25 minutes to 30 hours,"[21] which raises some question about the consistency of difficulty, familiarity, and intrinsic interest inherent in these tasks. Since the interaction data from these tasks were later collapsed together for the purpose of statistical analysis, there is some doubt as to the appropriateness of paying such little attention to the nature of the tasks, particularly in light of Fisher's expressed interest in "in-depth research of task behavior".[22]

There was considerable conceptual confusion over the task "types" employed in these eight studies. None of the investigators employed more than one type of task. None of them tested for task effects or differences, either. Leathers and Larson contend that their groups discussed "problem-solving" tasks. Fisher refers to the tasks discussed in his groups as "decision-making". A "ranking task" was discussed by Stech's groups. McCroskey and Wright, Bostrom, and Burgoon, each conceived of their task assignments as "discussion tasks", although they varied from "a discussion on" to "a topic assignment" to a "propositional discussion".

Several important weaknesses are evident in these task definitions. First, there is no clear distinction made between the different types of tasks being used. Except for the formal restriction offered by Steeh, it is implicitly suggested that the results of these studies can be extrapolated to task-oriented groups in general. This raises the question of whether data collected in groups assigned decision-making or problem-solving tasks is comparable to data collected in groups assigned discussion tasks. The experimental literature suggests that they are not comparable. Hackman has produced the clearest conceptualization of task differences.[23] He contends that tasks can be differentiated on the basis of "the kinds of cognitive materials with which a group works."[24] In this respect, discussion tasks differ markedly from problem-solving tasks. Discussion tasks require groups to work with issues, while problem-solving tasks require groups to deal with proposed actions or implementations. According to Hackman the differences between these two types of tasks can be reduced to the difference between asking "Should..." and/or "What changes, if any..." (discussion tasks), and "How can..." (problem-solving tasks).[25] Four studies testing this conceptualization have been reported in the literature.

Hackman found that task type determined up to 50 per cent of the variance on group output dimensions.[26] Morris discovered that different task types required different levels and categories of interaction.[27] Kent and McGrath replicated Hackman's experiment and reported that task type was a better predictor of performance than group characteristics.[28] And Hackman and Vidmar found substantial task type effects regardless of group size.[29] Together, these studies indicate that some specific limitations should be placed on the kinds of groups to which the results of at least six of the sampled studies can be generalized.

In three of these studies the task is inappropriately used within the experimental design. On the basis of the Illinois task type research, Hackman concluded that "unless tasks are appropriately held constant, counterbalanced, or sampled throughout an experimental design, a real possibility exists that the results of a study may be seriously confounded with unintended task effects."[30] In the Fisher, Bostrom, and Larson reports there is no indication that the tasks have been held constant, counterbalanced, or adequately sampled. No estimate of task effects is given. The potential for error variance attributable to task differences exceeds minimum confidence levels and the stated conslusions of these studies may be seriously questioned.

Stech is the only investigator who qualifies his results in terms of the task employed. He restricts generalization to "one kind of task" and he proposes that "some types of tasks undoubtedly lead to much higher degrees of structure than others."[31] While Stech should be commended for carefully qualifying his results, his statements bring up an additional question. How representative of ranking tasks in general was the ranking task used by Stech? Shaw's research with "task dimensions" has demonstrated that tasks can be very similar in what they ask the group to do, e.g., ranking cities according to their population, solving arithmetic problems, creating poems, etc., and at the same time be very dissimilar in important a priori task dimensions such as difficulty, familiarity, clarity, interest, intellectual requirements, and cooperation requirements.[32] In terms of Stech's research, a plausible rival hypothesis appears when these task qua task properties are considered. Any one of them might have contributed as much or more to interaction structure than the act of ranking fourteen types of music did. Stech's extrapolation to "one type of task" would have been more appropriate if a more complete sample of ranking tasks had been used or if he had pro-

vided evidence that the dimensions of his ranking task were representative of the dimensions of ranking tasks in general. Without such evidence, the implications should be restricted to one task rather than one kind of task.

Six of the other seven investigators also disregarded a priori task dimensions. Only Burgoon considered a task qua task property; he pretested his discussion topic for familiarity.

Did the tasks used in these "task-oriented" small group studies produce an unbiased estimate of experimental effects? This inspection of the types of tasks used and the ways in which they were used reveals that:

(1) There was little similarity between investigators in terms of the kinds of tasks assigned. One study used a ranking task, while the others used either problem-solving or discussion tasks.

(2) There was considerable confusion over task type. Despite the current interest in theory, none of the investigators provided a theoretical justification for their task choice. Only one researcher qualified results on the basis of task type. There was an implicit suggestion in each of the other studies that the results could be generalized to any task-oriented group.

(3) Three studies may have been seriously confounded by their use of tasks within the experimental design. In these studies, the tasks were neither uniform, counterbalanced, or sampled.

(4) Only one study measured any a priori dimensions of the group task and this study considered only one dimension, familiarity.

It appears that small group researchers in speech communication either patently disregard the task variable or are unaware of its importance. Such inattention seriously compromises the internal validity of these studies.[33]

Reliability Measurement

This report deals directly with the issues surrounding the degree of confidence which can be placed in small group research. In studies which utilize observer measures of interaction, confidence depends at least partially on the reliability of the measuring instrument. Six of the eight sampled studies based their conclusions on observer measures of communication behavior. One of the other studies was designed to produce a new measure of interaction behavior. Table II presents a summary of the reliability reports which appeared in these seven studies. Included are the dependent variable, the number of coders, the reliability estimate, and the method of computation.

- -

Insert Table II about here

- -

Reports of coding reliability are noticeably incomplete in all of these studies, with the possible exception of McCroskey and Wright's. In terms of the measuring instruments used, the seven studies can be separated into two major classes: those which use non-parametric category systems (Fisher, Stech, Bostrom, and Larson), and those which use parametric rating scales (two by Leathers and McCroskey and Wright's).

The four "category" studies are exceedingly neglectful in their treatment of reliability. Fisher, for example, reports that he made "several checks" for "agreement" among "minimally trained coders" and that these "checks" yielded agreement beyond .86. The reader is not told how many coders were used, whether the data are marginal totals or act-by-act comparisons, whether the reported coefficient represented a total realiability or an item reliability estimate, or what statistic was used.

Stech's presentation is more complete, but even less satisfactory. In their appraisal of interaction scoring and reliability problems, Waxler and Mishler present evidence that scoring procedures substantially affect score distributions. Their data suggest that studies of sequential interaction should use act-by-act measures of reliability.[34] In Stech's study the reader is deprived of the information necessary to assess the validity of the findings. Stech reports that a "reliability check" was made between the investigator and one other coder on one of the five transcripts, but that all of the coding was done by the investigator. Several questions are raised by these procedures: Why was only one of the five transcripts subjected to a reliability test? How did the investigator decide which of the five transcripts would be tested? Why was a rank-order correlation computed, since the category system did not require ranking? Why was all the coding done by the investigator? Why wasn't an act-by-act agreement coefficient computed?

Bostrom apparently feels that one person, working alone behind a one-way mirror, can accurately record the frequency and direction of verbalizations. Since only one observer was used, no agreement estimate could be computed. In previous research, observer seating position and angle of vision have been shown to affect coder judgment. If readers are expected to take Bostrom's concluding indictment of the advice offered in contemporary group discussion texts seriously, then some precedence or rationale for using only one observer should have been provided.

A footnote in Larson's article informs us that "the Berg instrument was pretested with three independent coders for reliability," and that "theme type and identification of theme initiator exceeded 85%."[35] Larson reports that three separate tests were made; apparently the three raters were compared

to one another two at a time. Again the reader is not informed of whether the data were marginal or act-by-act distributions. It is known that raters made several thousand judgments. It is also known that the Berg instrument contains only a few categories. Given this information, an act-by-act estimate would be the preferable measure of agreement, since marginal proportions could represent markedly different sets of acts.

Leathers' two studies and the McCroskey and Wright investigation employed very similar instruments. In fact, nine of McCroskey and Wright's thirty scales were taken from Leathers' LFRI. Although raters in both studies used similar procedures, making interval scale judgments of individual verbal messages, the investigators utilized very different methods of reliability estimate. McCroskey and Wright applied Hoyt's analysis of variance method to their data, while Leathers computed product moment correlations in both of his studies. The methodological differences may reflect a distinctive difference between the two investigations. The pearson coefficient is most often considered to yield a measure of inter-observer agreement, while Hoyt's test yields a measure of instrument reliability. Measures of user agreement are not comparable to measures of test reliability. Therefore, there is some doubt about the legitimacy of comparing the results of these studies.

Concerning the overall reliability of the measures of interaction behavior in these studies, the following is concluded:

(1) Reports of reliability are not a crucial consideration in the publication of small group research. Only one study gave more than minimal attention to estimating observer agreement or instrument reliability. Four of the studies failed to report the method of computation employed, even in footnote.

(2) Investigators who use category systems do not report the types of comparisons for which agreement estimates are provided. There is a substantial difference between category-by-category correlations and act-by-act correlations. It was not clear which method was being used by these researchers.

(3) Although investigators usually report high levels of agreement among coders, there are few reports of high instrument reliability. In only one case is it made clear that more than "a few" coders have used the instrument. Without such evidence, the reliability of the instrument remains in doubt.

Discussion

As a group, these eight studies do not substantiate the claims that small group research is methodologically rigorous and sophisticated. Examples of failure to report instrument reliability and independent support of validity suggest instead a relative devaluation of measurement as a problem preliminary to research.[36] Arguing whether method is primary or secondary to theory or whether the emphasis on theory building has resulted in a harmful neglect of method seems somewhat less important than demanding that published studies produce unbiased estimates of experimental effects. Without such a standard, efforts to build theory from research will be fruitless.

In light of the methodological problems uncovered in this review, the following recommendations seem appropriate:

(1) The task variable is always a stimulus in small group research. Investigators can decrease the possibility of unwanted error by taking greater care in selecting and using tasks. If only one task is used, generalization should be highly restricted. If multiple tasks are used, they should be counterbalanced and the possibility of task differences should be measured. Task qua

task properties should also be measured, since they may interact with independent variables under investigation. This extreme care is necessary because, as Hoffman warns, "the common practice of not identifying the problem used...deprives the researcher of a great deal of important and necessary information for his understanding of the experiment reported."[37]

(2) Reliability reports should be more complete and more accurate. Presently, almost any means of computing reliability or agreement appear to be acceptable. This should not be true. Both act-by-act and category-by-category coefficients should be reported, so that readers can make more accurate judgments about the validity of the results. Waxler and Mishler point out the necessity and advantages to more stringent reliability requirements when they conclude that "for every system the effect of the scoring procedure on the distribution of scores must be considered as well as the reliability of the instrument...knowledge about those aspects of the instrument will allow a more accurate interpretation of the data obtained in one study and more useful comparisons of several studies employing the same instrument."[38]

(3) Should theory be given precedence over method? Regardless of what position is defended, it is likely that researchers will continue to hold a wide range of values running the gamut from emphasis on measurement to emphasis on theory. Since "all theorizing is based on some sort of data and all data gathering involves some direction if only by implicit theories,"[39] it would be most advantageous to at least recognize the interdependent nature of the two.

Table I: Flow Sheet of Sample Small Group Research Published in Speech Monographs 1970-71.

Author - Issue	Purpose	Task	Conclusions
Fisher, B. Aubrey, Speech Monographs, March 1970.	"...to discover the nature of the interaction process across time leading to group concensus on decision-making tasks" (p. 54.) "...to discover patterns of verbal task behavior" (p. 54).	No specific task information given - Groups studied were similar in that: (1) Group goal was to achieve concensus; (2) Group successfully accomplished its task. However, time required to solve the decision making tasks varied from 25 min. to 30 hours.	"This study demonstrates that the interaction patterns of task behavior follow a consistent pattern of progression across time as groups make decisions..." (p. 65).
Leathers, Dale G., Speech Monographs, August 1970.	"...to measure the effects of...trust destruction on the small group communication process" (p. 181).	(1) Ss played money game with a partner in which they were told that winning was based on their ability to communicate. (2) Ss participated in 1 hr. discussion of "what action(s) if any, should be taken by UCLA to better identify with the Black Student Union." (3) Ss were given 1 week to prepare for discussion.	"The results...provide empirical support for the need to build trust among discussants...The discussant may be well advised to pay as much attention to the trust building potential of his message as to such traditional performance standards as acceptable evidence and sound reasoning" (p. 187).
Stech, Ernest L., Speech Monographs, November 1970.	"...to investigate the degree of structure in one kind of discussion task...to detect the degree of distributional and sequential structure in group discussion concerning a ranking task" (p. 249).	"...to rank fourteen types of music in the order in which the average American factory worker would prefer them" (p. 249).	"...discussion groups working on at least one kind of task do exhibit a fair degree of predictability." "Some types of tasks undoubtedly lead to much higher degrees of structure than others..." "The data indicate that there is some degree of... structure in interaction" (p. 255).
Bostrom, Robert N., Speech Monographs, November 1970	To describe group communicative patterns by: (1) Comparing theoretical possibilities to actual behavior; (2) Find the communicative act frequencies; (3) Determine the relationship between communicative patterns and other group factors.	"Each group was assigned a topic and after a twelve minute preparation period, discussed the topic for forty minutes" (p. 254).	"The high sender tends to send more than he receives...and to send more to a few persons rather than the group as a whole" (p. 261). "The individual who sends more than he receives is... more satisfied with the discussion than the member who receives more than he sends" (p. 263).
Burgoon, Michael, Speech Monographs, June 1971.	To test the proposition that "people with different levels of tolerance for ambiguity react differently to having to process conflicting information to arrive at group concensus" (p. 121).	A 30 minute discussion of the proposition: "The current civil defense program should be overhauled." Task was chosen on the basis that (1) Ss were relatively unfamiliar with it, and (2) Issue was controversial enough to insure discussion.	"...amount of existing conflicting information will mediate evaluation of the task" (p. 124).
Larson, Charles V., Speech Monographs, August 1971.	To compare the verbal behavior of initially leaderless, task-oriented groups in which a leader has clearly emerged with the same type of groups, in which no leader has emerged.	"All groups had similar task assignments: They were asked to prepare and present a written and oral report which focused on a problem which concerned them" (p. 178).	"...when a leader (emerges in a group), the overall group attention span is extended and the group is able to concentrate its discussion on single ideas for longer periods of time" (pp. 180-181).
Leathers, Dale G., Speech Monographs, August 1971.	To test for the direction qualities and degree of disruptiveness of feedback associated with a specific kind of message variable.	Refers to discussion in pilot work as "problem-solving discussions" (p. 182). Ss had two weeks to prepare for an 80 minute discussion on: "What actions, if any, should UCLA take to work closely with the Reagan Administration?"	"The confirmed relationships between specific levels of abstraction and facetious and specific feedback qualities in problem-solving discussions are determinant interactions" (p. 189).
McCroskey, James C., and Wright, David W., Speech Monographs, November 1971.	To develop an instrument for measuring small group communication which would allow parametric statistical analysis.	A 30 minute discussion on what should the university do about parking in the campus area?	"The I.B.M. ...can be used by evaluators...with the expectation that the factor structure in the resulting data will be essentially the same" (p. 339).

Table II: Reliability Estimates and Methods of Computation for Eight Selected Small Group Experiments

Author	Dependent Variable	No. of Coders	Reported Reliability	Method of Computation
Fisher	Fisher's Original Category System.	Not given. Used "minimally trained coders" (p. 55).	> .86 (agreement).	"several checks" (p. 55).
Leathers (1970)	Leather's Feedback Rating Instrument.	Two "carefully trained coders" (p. 185).	From .81 to .89 on the 9 scales.	Pearons Product-Moment.
Stech	Category system specifically designed for the ranking task used.	One; the investigator.	.87 on 1 transcript (of the 5 used in the study).	Rank Order correlations between investigator and 1 other coder.
Bostrom	Frequency and direction for each verbalization for each discussant.	One.	None.	None.
Larson	Berg's Time-Devoted-To-Themes Instrument	Three	> .85 for three separate tests.	Not given (Pearson assumed).
Leathers (1971)	LFRI	Two "carefully trained judges" (p. 186).	Exceeded previously established criteria (Leathers, 1970).	Pearson r
McCroskey and Wright	Interaction Behavior Measure IBM	Thirteen inexperienced raters. Replicated by twelve graduate student raters.	.64 to .92 for each scale.	Hoyt and Guilford's ANOV Procedure.

Footnotes

1. J.E. McGrath and I. Altman, Small Group Research: A Synthesis and Critique of the Field (New York: Holt, Rinehart, and Winston, 1966).

2. Ernest G. Bormann, "The Paradox and Promise of Small Group Research," Speech Monographs, 37, 1970, 211-216.

3. Dennis Gouran, "Response to 'The Paradox and Promise of Small Group Research'," Speech Monographs, 37, 1970, 217-218.

4. C. David Mortenson, "The Status of Small Group Research," Quarterly Journal of Speech, 56, 1970, 304-309.

5. Carl E. Larson, "Speech Communication Research on Small Groups," Speech Teacher, 20, 1971, 89-107

6. B. Aubrey Fisher, "Communication Research and the Task Oriented Group," Journal of Communication, 21, 1971, 136-149

7. Robert S. Cathcart, "Research in Small Group Communication: A Response," a paper presented at 1970 SCA convention, p. 1.

8. Bormann, 211.

9. Mortenson, 305-306.

10. Gouran, 218.

11. McGrath and Altman, 87.

12. McGrath and Altman, 88.

13. D. Byrne, "Assessing Personality Variables and Their Alteration," in P. Worchel and D. Byrne (Eds.), Personality Change (New York: John Wiley and Sons, 1964), 57.

14. Saul M. Sells (Ed.), The Stimulus Determinants of Behavior (New York: Ronald Press, 1963).

15. See for example: Richard Hackman, "Toward Understanding the Role of Tasks in Behavioral Research," Acta Psychologica, 39, 1969, 97-128; R. Heslin, "Predicting Group Task Effectiveness from Member Characteristics," Psychological Bulletin, 62, 1964, 248-256; and L. R. Hoffman and G. G. Smith, "Some Factors Affecting the Behavior of Problem-Solving Groups," Sociometry, 23, 1960, 273-291.

16. Marvin E. Shaw, "Social Psychology and Group Process," in J. B. Sidowski (Ed.), Experimental Methods and Instrumentation in Psychology (New York: McGraw-Hill, 1966).

17. Karl Weik, "Laboratory Experimentation with Organizations," in J. C. March (Ed.), Handbook of Organizations, (Chicago: Rand McNally, 1965).

18. These are: Edgar Borgatta, "Some Task Factors in Social Interactions," Sociology and Social Research, 48, 1963, 5-12; C. Morris, "Task Effects on Group Interaction," Journal of Personality and Social Psychology, 4, 1966, 545-554; and Bruce Tuckman, "Group Composition and Group Performance of Structured and Unstructured Tasks," Journal of Experimental Social Psychology, 3, 1967, 25-40.

19. Morris, 545-554.

20. These eight studies comprised a complete sample of all of the original small group research published in Speech Monographs, during 1970-1971. They were: B. Aubrey Fisher, "Decision Emergence: Phases in Group Decision Making," 37, 53-66; Dale G. Leathers, "The Process Effects of Trust-Destroying Behavior in the Small Group," 37, 180-187; Ernest L. Stech, "An Analysis of Interaction Structure in the Discussion of a Ranking Task," 37, 249-256; Robert N. Bostrom, "Patterns of Communicative Interaction in Small Groups," 37, 257-268; Michal Burgoon, "Amount of Conflicting Information in a Group Discussion and Tolerance for Ambiguity as Predictors of Task Attractiveness," 38, 121-124; Charles W. Larson, "The Verbal Response of Groups to the Absence or Presence of Leadership," 38, 177-181; Dale G. Leathers, "Testing for Determinant Interactions in the Small Group Communication Process," 38, 182-189; and James M. McCroskey and David W. Wright, "The Development of an Instrument for Measuring Interaction Behavior in Small Groups," 38, 335-340.

21. Fisher, 54.

22. Fisher, 53.

23. Richard Hackman, "Effects of Task Characteristics on Group Products," Journal of Experimental Social Psychology, 4, 1968, 162-187.

24. Hackman, 164.

25. Hackman, 164.

26. Hackman, 162-187.

27. Morris, 545-554.

28. R. N. Kent and J. E. McGrath, "Task and Group Characteristics as Factors Influencing Group Performance," Journal of Experimental Social Psychology, 5, 1969, 429-440.

29. J. R. Hackman and N. Vidmar, "Effects of Size and Task Type on Group Performance and Member Reactions," Sociometry, 33, 1970, 37-54.

30. Hackman, 183.

31. Stech, 255.

32. Marvin E. Shaw, Scaling Group Tasks: A Method for Dimensional Analysis, Technical Report No. 1, ONR Contract NR 170-266, Nohr-580 (11), University of Florida.

33. A similar point is made by Hackman, 1968, 183.

34. Nancy E. Waxler and Elliot G. Mishler, "Scoring and Reliability Problems in Interaction Process Analysis: A Methodological Note," Sociometry, 29, 1966, 28-40.

35. Larson, 178.

36. With the exception of the McCroskey and Wright Study.

37. Hoffman and Smith, p. 286.

38. Waxler and Mishler, p. 40.

39. Byrne, 57.